The role of the energy gap in protein folding dynamics

E.Pitard * and H.Orland
Service de Physique Théorique
Centre d'Etudes de Saclay, CEA, 91191 Gif-sur-Yvette
France
(February 1, 2008)

The dynamics of folding of proteins is studied by means of a phenomenological master equation. The energy distribution is taken as a truncated exponential for the misfolded states plus a native state sitting below the continuum. The influence of the gap on the folding dynamics is studied, for various models of the transition probabilities between the different states of the protein. We show that for certain models, the relaxation to the native state is accelerated by increasing the gap, whereas for others it is slowed down .

I. INTRODUCTION

Proteins are known to fold into a unique native state that is biologically active [1]. Despite their complexity and frustrated character, their folding rate is fast and the folding times vary from milliseconds to seconds. However, these times are long compared to the usual collapse time of a homopolymer (a few microseconds) [2]. This means that the diversity of the aminoacids and the resulting frustration play a crucial role in the slowing down. The main picture that has emerged to describe this type of dynamics is that of a "funnel-like" phase space [3]: the folding pathway of a protein in phase space is along a unique but rough funnel towards its ground state. However, so far, this picture is not based on microscopic models.

Numerical simulations that have been dealing with lattice models of short disordered polymers tend to support this funnel picture. A number of these simulations seem to show that "good folders" (sequences that fold rapidly to their ground state, and are thus good candidates to represent real proteins) indeed follow a "funnel" in the energy landscape, whereas "bad folders" possess a large number of energetic traps along the dynamical path that slow down the folding process [4,5].

Several studies have tried to relate the foldability of model proteins with their energy gap. It has been argued in ref. [6] that model proteins which have a large energy gap between the native state and the first conformationally different (compact) state fold rapidly. On the other hand, other studies [7] have shown that the parameter which governs the rapidity of folding is the distance of the folding temperature to the collapse temperature.

The aim of this paper is to show that the situation is not so simple, and that in particular, the folding rate depends very much on the dynamics used for the simulations. In particular, we show that for certain transition rates, a large gap accelerates the folding, whereas for other models, it may slow it down.

II. MODELIZATION BY A TRUNCATED REM

Many models have stressed the analogy between protein folding and the thermodynamics of heteropolymers [8]. In a mean field approach, some models of quenched disordered polymers are similar to a Random Energy Model (REM) [9,10]; the low lying states of the protein are identified with the low-lying states of the REM, which are responsible for the slow dynamics. Although this analogy is questionable as far as real proteins are concerned [11,12], we will adopt this framework for the dynamical models studied in the following.

For the REM as well as for the Sherrington Kirkpatrick (SK) model [13], it has been shown that i) all low lying states have the same extensive energy and ii) the distribution of the corrections to extensivity of their free energies is exponential (bound at high energy but not at low energy).

For proteins, the energies are bound below by that of the native state and above by that of a swollen coil. In addition, the energy landscape is known to be very rugged and the number of misfolded states (at fixed energy E) is known to grow very rapidly with E. It is thus natural, by analogy with many disordered systems, to assume an exponential distribution for the energies of the lowest lying misfolded states and to isolate the native state below the continuum of energies.

In our model, we describe the phase space of the protein as consisting of M misfolded states E_α with $\alpha = 1, \ldots, M$ and one native state E_0. The distribution of energies of the misfolded states is continuous, given by:

$$p(E) = \beta_c e^{\beta_c (E - E_c)}$$

where β_c is a parameter (related to the glass transition temperature of the REM) and E_c is the energy of the

*e-mail:pitard@spht.saclay.cea.fr

highest state.

More precisely, the energy levels are such that $E_{min} \leq E \leq E_c$, where E_{min} is the energy of the first misfolded state. Therefore, Boltzmann weights $B_\alpha = e^{-\beta E_\alpha}$ vary between $B_{min} = e^{-\beta E_c}$ and $B_{max} = e^{-\beta E_{min}}$. To account for the native state, we include an additional state with energy E_0 and Boltzmann weight B_0.

The energy gap Δ is defined by $\Delta = E_{min} - E_0 > 0$.

The main consequence of the truncation of the energy distribution is that the system will never spend a very large time in one of the energy traps, and that $P_\alpha(t)$ will always relax towards its equilibrium value P_α^{eq} with a finite relaxation time at large times. Slow dynamics features such as aging won't appear either; this is in contrast with the case when the distribution is not truncated, which has been studied extensively recently [14].

The calculations are made by taking first the limit $M \to \infty$ before taking the limit of large times. This is justified if the number of metastable conformations is large enough.

The dynamics is based on the master equation:

$$\frac{dP_\alpha}{dt} = \sum_\beta W_{\alpha\beta} P_\beta(t) - \sum_\beta W_{\beta\alpha} P_\alpha(t) \qquad (1)$$

where $P_\alpha(t)$ is the probability of occupation of the energy level E_α at time t , and $W_{\alpha\beta}$ is the transition rate from the energy level E_β to the energy level E_α. In all the cases we have studied, detailed balance is satisfied:

$$\frac{W_{\alpha\beta}}{W_{\beta\alpha}} = e^{-\beta(E_\alpha - E_\beta)} \qquad (2)$$

When solving the master equation, the quantities that are usually calculated are averages over the distribution of disorder $p(E)$. This is justified in the case of a macroscopic system with short range interactions, but not in the case of a protein, which is too small an object for self-averageness to hold.

This is why, in contrast with other studies, we have calculated quantities which are not averaged over the distribution of disorder. We considered three cases, depending on the choice of transition rates between energy levels. We find that if the transitions rates depend only on the final state, the relaxation is accelerated by a large gap whereas if they only depend on the initial state, the dynamics is slowed down; in the intermediate case, the situation is more complex.

III. RESULTS

In this section, we present the results of the calculations, which will be detailed elsewhere.

Studies as the ones presented below have already been made for a non truncated distribution of energies: disorder-averaged quantities show stretched exponential or power law behavior at large times [15,16]. The same models have been used in ref. [17] in the context of heteropolymer folding.

A. Case where the transition rates depend only on the final state: $W_{\alpha\beta} = e^{-\beta E_\alpha}$

If E_0 and a gap are included in the analysis, the master equation can easily be solved, leading to an exponential behavior, $P_\alpha(t) = \frac{e^{-\beta E_\alpha}}{Z} + (P_\alpha(0) - \frac{e^{-\beta E_\alpha}}{Z}) e^{-Zt}$, where $Z = \sum_\alpha B_\alpha$. Since $Z \simeq B_0 + M\bar{B_\alpha} = B_0 + \frac{x}{1-x} B_{max}$ (the bar denotes the average over the distribution of energies), where $B_0 = B_{max} e^{\beta\Delta}$, the relaxation time is $\tau = \frac{1}{B_{max}(e^{\beta\Delta} + \frac{x}{1-x})}$.

Let us compare the dynamics of two protein sequences that differ only by their native energies E_0 keeping E_{min} (or B_{max}) constant. As seen above, the relaxation time τ decreases when the gap Δ increases. Indeed, in such a dynamical scheme, jumps towards the native state are enhanced and the lower its energy, the faster it is populated. The dynamics is illustrated by Fig.1 where the probability of occupation of the native state $P_0(t)$ is plotted for different values of the gap. On Fig.2, we show the relaxation time as a function of the gap.

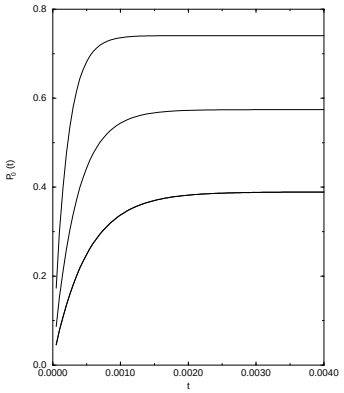

FIG. 1. $P_0(t)$ in the case where $W_{\alpha\beta} = e^{-\beta E_\alpha}$ for three sequences with the same distribution of energies but different gaps. From top to bottom, $\Delta = 1.5$, $\Delta = 1$, $\Delta = 0.5$, $\beta = 1.5$ and $M = 100$. The dynamics is faster for larger values of the gap.

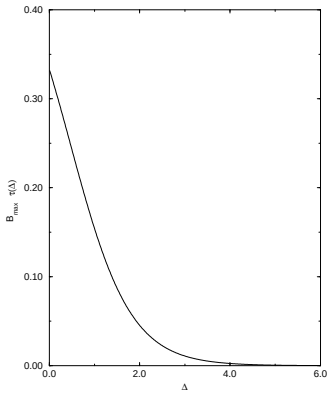

FIG. 2. Relaxation time $\tau(\Delta)$ as a function of the gap Δ in the case where $W_{\alpha\beta} = e^{-\beta E_\alpha}$, with $\beta = 1.5$ and $\beta_c = 1$.

B. Case where the transition rates only depend on the initial state: $W_{\alpha\beta} = \frac{1}{M} e^{\beta E_\beta}$.

As was noted in ref. [17], this corresponds to the case of a unique barrier at energy E^*, through which the system must pass in order to make a transition from state α to β.

Calculations can be made in both cases where the initial conditions are uniform (all states are equally populated) or delta-like (the initial probability of occupation of a specific state is one). They lead to the same conclusions concerning the relaxation times, and we give here the results for the case of uniform initial conditions. At short times, the probability of occupation of states follows a power-law:

$$P_\alpha(t) \simeq \frac{1}{x B_{max}^x} \frac{1}{\Gamma(1-x)\Gamma(x)\Gamma(1+x)} t^x$$

At longer times, the relaxation is exponential. For a non-native state $\alpha \neq 0$,

$$P_\alpha(t) \simeq \frac{1-x}{x} \frac{1}{B_{max}^2 e^{\beta\Delta}} \left(\frac{1}{ab} + \frac{1}{a(a-b)} e^{-at} - \frac{1}{b(a-b)} e^{-bt} \right)$$

and for the native state,

$$P_0(t) \simeq \frac{1}{1 + \frac{x}{1-x} e^{-\beta\Delta}} (1 - e^{-bt})$$

These expressions involve two time constants: $\tau_\alpha = \frac{1}{a} = e^{-\beta E_\alpha}$, which is the relaxation time of the energy level α in the absence of a gap, and $\tau_0 = \frac{1}{b} = \frac{B_{max}}{\frac{1-x}{x} + e^{-\beta\Delta}}$. This last relaxation time is the one that governs the long time dynamics of the native state.

Contrarily to the previous case, keeping E_{min} (or B_{max}) fixed, τ_0 now increases as Δ increases. In this scheme, the energy landscape can be viewed as a collection of energy traps; as the energy E_0 decreases, it becomes more difficult to escape from this state; low energy states are populated quite rapidly, and the system remains trapped in the native state; as a result, the relaxation to the Boltzmann distribution is slowed down. This effect is more pronounced as the energy gap gets bigger (see Fig.3 and 4).

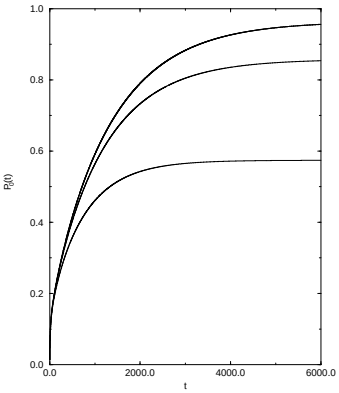

FIG. 3. $P_0(t)$ in the case where $W_{\alpha\beta} = e^{\beta E_\beta}$ for three sequences with the same distribution of energies but different gaps. From top to bottom, $\Delta = 3$, $\Delta = 2$, $\Delta = 1$. $\beta = 1.5$ and $M = 100$. The dynamics slows down for large values of the gap.

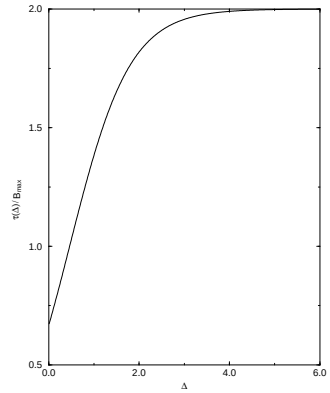

FIG. 4. Relaxation time $\tau(\Delta)$ as a function of the gap Δ in the case where $W_{\alpha\beta} = e^{\beta E_\beta}$, with $\beta = 1.5$ and $\beta_c = 1$.

C. Intermediate case: $W_{\alpha\beta} = e^{-\beta((1-\lambda)E_\alpha - \lambda E_\beta)}$

In this case, the transition rates depend both on the initial and final state through the above formula, where λ is a parameter between 0 and 1. Relaxation to the native state is described by a new relaxation time τ_0:

$$\tau_0 = \frac{B_{max}^\lambda e^{\lambda\beta\Delta}}{Z} \frac{1}{1 + \frac{1-x}{x} e^{\beta\Delta}}$$

where $Z = \sum_\alpha B_\alpha^{1-\lambda}$.

The dependence of τ_0 on the gap is now slightly more complicated. One can distinguish two regimes:

1. if $\lambda \ll 1-x$ and $\lambda \leq \frac{1}{2}$, τ_0 decreases as Δ increases (see Fig. 5 and 6), and the relaxation is accelerated. On the other end, if $\lambda > \frac{1}{2}$, τ_0 increases with Δ for $\Delta < \Delta_0$ and decreases with Δ for $\Delta > \Delta_0$, where $\beta\Delta_0 = \log(\frac{2\lambda-1}{2(1-\lambda)}\frac{x}{1-x})$.

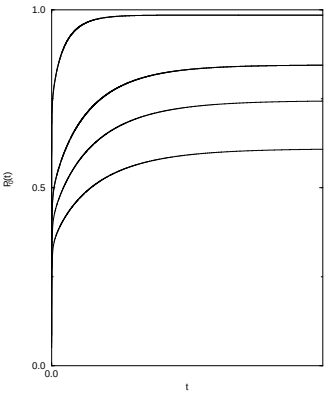

FIG. 5. $P_0(t)$ in the case where $W_{\alpha\beta} = e^{-\beta((1-\lambda)E_\alpha - \lambda E_\beta)}$ for four sequences with the same distribution of energies but different gaps. From top to bottom, $\Delta = 2$, $\Delta = 1$, $\Delta = 0.75$, $\Delta = 0.5$. We have chosen $\lambda = 0.5$, $\beta = 2.5$ and $M = 100$. For this set of parameters, the dynamics is faster as the gap increases.

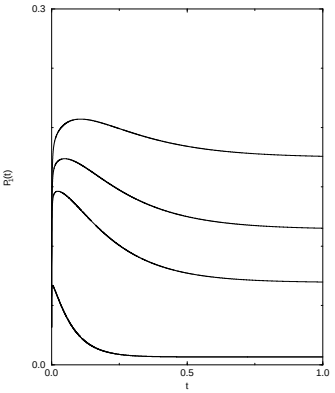

FIG. 6. $P_1(t)$ in the case where $W_{\alpha\beta} = e^{-\beta((1-\lambda)E_\alpha - \lambda E_\beta)}$ for four sequences with the same distribution of energies but different gaps. From top to bottom, $\Delta = 0.5$, $\Delta = 0.75$, $\Delta = 1$, $\Delta = 2$. We have chosen $\lambda = 0.5$, $\beta = 2.5$ and $M = 100$.

2. if $\lambda \gg 1-x$, τ_0 is an increasing function of Δ for $\Delta < \Delta_0$ and a decreasing function of Δ for $\Delta > \Delta_0$, where $\beta\Delta_0 = \log(\frac{x\lambda}{(1-x)(1-\lambda)})$.

IV. DISCUSSION

It seems difficult to compare our results with experiments in real proteins, since no systematic study of the gap of given sequences has been carried out experimentally. However, there has been a number of lattice simulations that lead to a variety of results and interpretations. Klimov and Thirumalai [7] claim that there seems to be no direct correlation between the gap and the folding dynamics.

On the other hand, Sali, Shakhnovich and Karplus [6] have shown that the "best" folders are those with the largest energy gap. In all cases, the simulations are performed on very short chains (27 monomers at best) and the interactions between monomers are random with a Gaussian distribution. The energy of a configuration is given by

$$E = \sum_{i<j} \Delta(r_i, r_j) B_{ij}$$

where $\Delta(r_i, r_j) = 1$ if monomers i and j are neighbors on the lattice and

$$P(B_{ij}) = \frac{1}{(2\pi B^2)^{\frac{1}{2}}} \exp\left(-\frac{(B_{ij}-B_0)^2}{2B^2}\right)$$

with a negative parameter B_0 in order to mimic the hydrophobic character of the solvent. These previous studies are applicable only to very short proteins. For longer chains, there are no results indicating how a protein will find its folding path through its complicated energy landscape. Some attempts have been made to understand analytically the dynamics of random heteropolymers [18] [19]; and at present, the debate regarding the interpretation of the simulations [7] [6] [4] [20] is still open.

Our phenomenological approach based on a REM energy landscape, although far from realistic, might capture the long time relaxation laws of random chains. The finiteness of objects such as proteins is taken into account by truncating the energy spectrum. Moreover, we show that the dynamics depends on the whole energy landscape, rather that only on one parameter, such as the gap.

According to the type of transition rates used in the dynamics, one can obtain different behaviors for the relaxation as a function of the gap. It seems difficult to decide which one is more realistic. It would be interesting to have systematic results on the dynamics of synthesized sequences, generated for example through mutation experiments [21].

[1] C.B. Anfinsen, *Science,* **181**, 223-230 (1973).

[2] P.G. de Gennes, *J. Phys. Lett.,* **46**, L639-L642 (1985).
[3] J.D. Bryngelson, J.N. Onuchic, N.D. Socci, P.G. Wolynes, *Proteins: Structure, Function and Genetics,* **21**, 167-195 (1995).
[4] R. Mélin, H. Li, N. Wingreen, C. Tang, preprint cond-mat/9806197.
[5] M. Cieplak, M. Henkel, J. Karbowski, J.R. Banavar, preprint cond-mat/9803019.
[6] A. Sali, E. Shakhnovich, M. Karplus, *Nature,* **369**, 248-251 (1994).
[7] D.K. Klimov, D. Thirumalai, *Phys. Rev. Lett.,* **76**, 4070-4073 (1996).
[8] T. Garel, H. Orland, E. Pitard, *Protein folding and heteropolymers,* in "Spin Glasses and Random Fields", A.P. Young (ed.), World Scientific, Singapore (1997).
[9] T. Garel, H. Orland *Europhys. Lett.,* **6**, 307-310 (1988).
[10] E.I. Shakhnovich, A.M. Gutin, *J. Phys. A.,* **22**, 1647-1659 (1989).
[11] A. Irback, C. Peterson, F. Potthast, LU TP preprint 95-31, October 1996.
[12] V.S. Pande, A.Y. Grosberg, C. Joerg, T. Tanaka, *Phys. Rev. Lett.,* **76**, 3987-3990 (1996).
[13] M. Mézard, G. Parisi, M.A. Virasoro, *J. Phys. Latt.,* **46**, L217-L222 (1985).
[14] C. Monthus, J-Ph. Bouchaud, *J. Phys. A: Math. Gen.,* **29**, 3847-3869 (1996).
[15] C. de Dominicis, H. Orland, F. Lainée, *J. Phys. (France),* **46**, L463-L466 (1985).
[16] G.J.M. Koper and H.J. Hilhorst, *Europhys. Lett.,* **3**, 1213-1217 (1987).
[17] E.I. Shakhnovich and A.M. Gutin, *Europhys. Lett.*, **9**, 569 (1989).
[18] J-R Roan, E. Shakhnovich, *Phys. Rev. E,* **54**, 5340-5357 (1996).
[19] D. Thirumalai, V. Ashwin, J.K. Bhattacharjee, *Phys. Rev. Lett.,* **77**, 5385-5388 (1996).
[20] A. Sali, E. Shakhnovich, M. Karplus, preprint cond-mat/9606037. D. Klimov, D. Thirumalai, preprint cond-mat/9607060.
[21] R.E. Burton, J.K. Myers, T.G. Oas, *Biochemistry,* **37**, 5337-5343 (1993).